
UCRL-TR-201447

# Recommendations for SZ/TSPA Model Uncertainty Analysis Concerning the Yucca Mountain Project

*K.T. Bogen*

**October 10, 2003**

Environmental Science Division
Energy and Environment Directorate

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

# Recommendations for SZ/TSPA Model Uncertainty Analysis Concerning the Yucca Mountain Project

by
Dr. Kenneth T. Bogen
LLNL (L-396, 925-422-0902, bogen@LLNL.gov)
October 10, 2003

This report considers the problem of how best to evaluate the stability (i.e., sampling reliability) of Monte Carlo outputs obtained for two Yucca Mountain Project (YMP) modeling components, namely, outputs for the Saturated Zone (SZ) model, and for the Total System Performance Assessment (TSPA) model. One approach considered is the one that has been employed to date, namely, the application of Monte Carlo methods. Also considered in this context are potential improvements that might be obtained by the additional use of a Monte Carlo "quitting rule", such as that defined by Woo (1991), to select the number of Monte Carlo sample runs to perform. By the Monte Carlo approach, each output-value sample (realization) is calculated as a function of a sample-value vector of stochastic realizations, each of which in turn corresponds to a value of a corresponding distributed input variable. Abstractions from the SZ model and the Biosphere model are both used as input to the TSPA model. Sets of stochastic realizations required for SZ and TSPA abstractions "expensive" to generate, so the practical issue addressed by a "quitting rule" is how to determine what number of realizations is "enough" for the purpose of characterizing sampling error in the Monte Carlo estimate obtained for a specified model output of concern. In the TSPA context, the model output of concern is generally considered the time evolution of the arithmetic mean value, an estimator of the expected value, of TSPA-generated annual dose $D(t)$ to the defined receptor within 10,000 y after waste-repository closure.

Recommendations below specifically address: (1) whether or not a Monte Carlo approach (such as one employing the Woo quitting rule) is an appropriate basis for undertaking SZ- or TSPA-related uncertainty analysis, and (2) what other method

might be more appropriate. While the following discussion and recommendations focus particularly on SZ applications, they also apply to other YMP model components including the TSPA. For this reason, some TSPA-specific issues are also addressed.

## 1. Are Monte Carlo Methods—e.g., Implemented Using the Woo (1991) Quitting Rule—an Appropriate Basis for Assessing Reliability of Estimated SZ or TSPA Model Output?

Monte Carlo methods are often used to characterize the output of complex probabilistic models. A key advantage offered by this standard approach is that it does not require an analytic expression of output value as a function of uncertain (distributed) input values. This approach is thus now commonly applied (e.g., using random or Latin Hypercube sampling methods to generate pseudo-random input-variable realizations) to characterize uncertainty in output of any complex model that is either impossible or too tedious to characterize using analytic methods (Helton and Davis, 2002). For this reason, Monte Carlo methods have also been used to date to characterize uncertainty YMP SZ and TSPA model outputs (BSC, 2001a-b).

Typical Monte Carlo analyses involve generating and sorting from $10^4$ to $10^5$ or more simulated output values for the purpose of characterizing output uncertainty. In the case of the SZ and (by virtue of its incorporation of the SZ model) the TSPA models, model-output realizations have been relatively expensive to obtain due primarily to the nonlinear structure and complexity of the SZ model, so typically far fewer than $10^4$ and frequently fewer than $10^3$ realizations have been generated for each characterization of SZ- or TSPA-model output uncertainties (BSC, 2001a). The following subsection (Section 1.1) explains why any characterization of SZ- or TSPA-model output uncertainties based on so few Monte Carlo realizations cannot be considered reasonably reliable *a priori*. The validity of this conclusion is independent from and unaffected by conditioning Monte Carlo sampling on the application of any Monte Carlo "quitting rule" (QR), such as that defined by Woo (1991), notwithstanding claims to the contrary asserted by Woo (1991). The QR defined by Woo (1991) and its questionable validity are discussed below in Section 1.2. Finally, Section 1.3 discusses the point that even if a mathematically valid QR (VQR) were correctly applied to one or more TSPA submodels (e.g., the SZ model) in order to obtain random sample values of submodel output to use as inputs for Monte Carlo TSPA evaluation, then the mathematical validity of a

subsequent VQR application to evaluate the TSPA itself (i.e., of any nested application of the VQR) would need to be established, and this issue has not previously been addressed (e.g., by Woo, 1991).

### 1.1. The "Curse of Dimensionality": Why Monte Carlo sampling cannot efficiently provide a reliable characterization of SZ/TSPA model uncertainty

The SZ and TSPA models each involve potentially complex (e.g., unspecified nonlinear) interactions among >20 distributed input variables (this point is discussed further in Section 2). Because it is possible that a combination of relatively (but not extremely) unlikely values of some (perhaps even a small) subset of these variables may produce a very large upward shift the value of modeled concentrations and/or exposures (and hence in the value of corresponding predicted risk), it simply is not possible to characterize uncertainty in estimated mean SZ or TSPA output with any reasonable degree of reliability using Monte Carlo methods ***unless*** sample size $N$ used is very large (e.g., $N >> 10^4$), i.e., far larger than sample sizes used for this purpose to date. This is true regardless of the Monte Carlo sampling technique used (e.g., Latin Hypercube vs. uniform sampling). The reason is simply the very large size of the sample space that must be undertaken to achieve such reliability, given the relatively large number $n$ of variables (and corresponding sampling dimensions) involved. This problem is referred to as the "curse of dimensionality" in Bayesian statistical literature, and it arises because the volume of sample hyperspace grows (hence its sample density shrinks) exponentially as a function of the sampling dimension $n$ (Bellman, 1961).

The following example illustrates why Monte Carlo methods do not efficiently yield reliable estimates of a poorly defined function of many distributed input variables. Suppose each of $n$ such suitably defined variables is known *a priori* only to be related monotonically to an output of interest (e.g., expected risk), and upper $p$-fractiles of subsets of $m$ (i.e., $m$-tuple subsets) of these variables could (in the absence of knowledge that this could not occur) potentially interact to greatly (e.g., super-multiplicatively) increase the value of corresponding risk conditional on those upper-bound input values, and hence similarly increase the true value of expected risk. It would follow in this case that any reasonably reliable assessment of uncertainty in estimated risk must be based on $s$ risk realizations that jointly reflect ≥1 sample of each $m$-tuple combination of upper $p$-fractiles. The chance that such specific combination

occurs is $p^m$, which clearly becomes very small quickly as $m$ increases, e.g., for values of $p \leq 20\%$. But for $m$ equal to just 2 or 3, the small likelihood $p^m$ might possibly be balanced by a large upward shift in corresponding predicted risk, which may in turn correspond to a non-negligible or even sizeable increase in expected risk. The key point is that, absent *a priori* knowledge about the true distribution of risk, there is simply no way of knowing that such a disproportionately increased risk level could be produced by a particular combination of $m$ upper-$p$ fractile values for $m$ corresponding inputs, without actually sampling each potentially relevant combination at least once.

Even if implemented using Latin Hypercube stratified sampling (LSH), Monte Carlo methods neither solve nor get around the dimensionality problem. In particular, Monte Carlo estimation done by LHS is unable to address this problem efficiently, and uniform random Monte Carlo sampling does even worse. For example, with $n = 20$, there are a total of $2^{20} \approx 10^6$ possible unique combinations of upper-vs.-lower bound values for all 20 variables. Sampling each of these combinations (by LSH or any other strategy) is clearly unfeasible in the YMP context. Of all possible combinations, only $n_m \doteq \binom{20}{m}$ potentially relevant $m$-tuple combinations of upper-bound values need be sampled in order to include each such upper-bound combination (UBC) at least once (e.g., $n_m = 190$, 1,140 and 4,845 for $m = 2$, 3 and 4, respectively). However using LHS (which guarantees that each sample will contain at least one among $n$ possible upper-bound values), each sample (among $N$ total samples) will contain a given number $n_{i,m} = \binom{i+1}{m}$ of $m$-tuple UBCs with a corresponding binomial probability $p_{i,} = \binom{n-1}{i} p^i (1-p)^{n-1-i}$, for $i = 1,..., 19$. For example, given $n \doteq 20$ and $p = 10\%$, the probability $p_0(m)$ that no $m$-tuple UBCs will be contained in particular sample is 13.5%, 42.0% and 70.5% with $m = 2$, 3 and 4, respectively. Finding the sample size $N$ sufficient to ensure a low probability $p_{cc}$ (e.g., $p_{cc} < 5\%$) that one or more potentially relevant UBCs remain unsampled is a doubly stochastic generalization of the classic "coupon collector" problem of determining an occupancy waiting time, namely: "What is the number $\mathcal{N}$ of store-visits (i.e., samples, assuming $m = 1$ coupon per sample) to ensure that among a total of $n_m$ different coupons, ≥1 remain uncollected with a probability that is *less than* $p_{cc}$." The well known

Poisson approximation is that $\mathcal{N} = N$ provided that $p_{cc} > [n_m(1-n_m^{-1})^N \approx n_m e^{-N/n_m}]$ and also that $p_{cc}$ is small, i.e., that

$$N > (n_m/m)[\mathrm{Log}_e(n_m/p_{cc})] \quad (1)$$

*conditional* on $m = 1$ (Feller, 1970). In our Monte Carlo problem involving $n = 20$ variables with $p = 10\%$, unique UBCs are the coupons of interest, but the number $n_{i,m}$ of possible $m$-tuple UBCs obtained per sample among all ($n_m$) possible unique UBCs varies randomly from sample to sample. The series expression (Johnson and Kotz, 1977) for the probability Prob($\mathcal{N} \le N$) conditional on any number $m$ of "coupons" ($1 \le m \le n$) was reported originally by Pólya in 1930. Evaluations of that expression indicate that Inequality (1) above provides a fairly accurate bound on $N$ in the general $m$-coupon case in all cases investigated (with $2 \le m \le 10$, $20 \le n_m \le 500$, $p_{cc} = 5\%$). In our doubly stochastic problem conditional on $m$ and $p_{cc}$, variations in $n_{i,m}$ will occur randomly over the $N$ required samples, so $N \approx \lceil \overline{N} / (1-p_0(m)) \rceil$, where $\lceil v \rceil$ = the greatest integer $\ge v$ and $\overline{N} = \sum_{i=m-1}^{n-1} \left\{ \left[ p_i/(1-p_0(m)) \right] \left[ (n_m/m) \log_e(n_m/p_{cc}) \right] \right\}$. These expressions yield corresponding minimum required values of sample size $N$ equal to 880, 11,559 and 126,634 with $m = 2$, 3 and 4, respectively. Thus if a number $N < 880$ of LHS samples were used in this example, there would be a significant probability that the joint upper-100$p$%-tail values of *at least one* of the 190 possible pairs of 20 variables would not be sampled at all, and consequently the Monte-Carlo estimate of output would not reflect the possibility that one or more unsampled upper-tail combinations might profoundly affect the true output of interest. Likewise, $>10^4$ realizations are required to reasonably ensure that all 3-variable joint UBCs are sampled and thus reflected in estimated output for this example—a sample size far greater than any previously used to characterize YMP-related uncertainties.

### 1.2. The validity of the Woo (1991) QR is questionable

The QR defined by Woo (1991) is based on the following definition of a "minimum necessary" number $W_N$ of Monte Carlo model runs, as a function of the current number

$N$ of such runs already completed, each of which runs yields a sample value $R_i$ of the model output $R$, for $i = 1,\ldots,N$:

$$W_N = (2e)^{-1} + [(e^{-1/2})\, g_N]/C \quad \text{for all} \quad g_N \geq \sqrt{\left(p^2 N - J_N\right)/J_N} \tag{2}$$

where:

$g_N$ = the sample coefficient of variation (i.e., the square root of the sample variance divided by $\overline{R}$, the sample arithmetic mean) of $R_i$ for $i = 1,\ldots,N$,

$J_N$ = the least number of the sampled risks $R_i$ that sum to a value $> pN\overline{R}$ with $p$ "close to 1" (e.g., $p \geq 0.95$ or $p > 0.90$),

$c = (R_{N+1} - \overline{R})/\overline{R}$ ,

$C$ = the maximum acceptable value of $c$, and

$e$ = Prob($C \leq c$) . (3)

An important aspect of this QR is the definition of $J_N$. According to Woo (1991, at p. 182), but inserting $J_N$ in place of Woo's notation $J(N)$ and bold font to key phrases:

> "In a sequence of Monte Carlo runs, many of **the sampled risk values** may be found to be zero of extremely small. In these circumstances, it is useful to **define the number of significant risks $J_N$ included within a sample of $N$ values**. A natural definition of $J_N$ is the least number of risks comprising more than 95% of the sum of **the risks obtained during a sequence of $N$ Monte Carlo runs**."

Note this definition refers not to "**any potential** sequence of $N$ Monte Carlo runs" but rather simply to "a sequence of $N$ Monte Carlo runs." In this context $R_i$ thus represents a particular set of $N$ realizations actually produced by $N$ corresponding Monte Carlo simulations—the same realizations to which the Woo QR is claimed to be applicable. It is also worth noting that $J_N$ defines a lower acceptable bound on $g_N = s_N / \overline{R}$, where (see Woo 1991 at p. 180) $\overline{R}$ = the sample (not population) mean of $R$, and $s_N$ = the square root of the corresponding sample (not population) variance. Thus, contrary to Woo's assertion (at p. 181, using the notation $V(N) = g_N$) that $g_N$ is "the coefficient of variation" of $R$, it is actually the *sample* coefficient of variation of $R_i$.

If Woo had defined $J_N$ and $g_N$, or just $J_N$, purely as functions of the true parent distribution of $R$, rather than in terms of the sample realizations $R_i$, then Woo's QR

would be a circular tautology with no practical application as a stopping rule applicable to any given set of $N$ sampled $R$-realizations, since in this case any particular estimate of $J_N$ (and hence the lower acceptable bound for $g_N$ required by the Woo QR) would be uncertain. If, on the other hand, $J_N$ were defined in terms of a true $R$-distribution well enough characterized to define $J_N$ with absolute certainty, then a sufficiently accurate estimate of the expected value of $R$, i.e., of $E(R) = E(\overline{R})$ to within $\pm 100(1-p)\%$ with $p = 1-e$, would by virtue of this definition of $J_N$ be provided directly by $\overline{R}$ conditional on $N = J_N$, without recourse to any QR.

The Woo (1991) QR thus asserts that using Eq. (2) to define $W_N$ guarantees that Eq. (3) is true, with $J_N$ and $g_N$ defined as functions of the samples $R_i$ and not of the parent distribution of $R$. Operationally, the Woo (1991) QR recommends terminating Monte Carlo sampling of $R$ as soon as (or reasonably soon after) $N > W_N$. Notably, Woo (1991) characterized this rule as a "general non-parametric rule for quitting Monte Carlo simulation," without reference to any *a priori* constraints (e.g., on the parent distribution from which $R$ is sampled) that are required in order to guarantee the truth of the rule.

The QR summarized above appears to offer an absolute and nonparametric bound on error of a sample mean, without specific conditions on the parent distribution from which samples are drawn. If valid, such a QR would comprise a fundamental statistical breakthrough, as important as Tchebychev's inequality, and would have direct applicability to complex estimation problems like those involved in YMP-related uncertainty analyses. However, the Woo (1991) paper was published not in a mathematical or statistical journal, where its underlying mathematical merits would likely have received rigorous peer review, but rather in a nuclear engineering journal. According to the Science Citation Index as of April 2003, the Woo (1991) paper had been cited only once, in a review paper by Helton (1993) that merely cited but did not specifically discuss the Woo (1991) paper.

The basis of the Woo (1991) QR is the following inequality, reported by Woo (1991), that is implied by a very similar (but stronger) inequality reported by Saw *et al*. (1984):

$$\text{Prob}\{|R_{N+1} - \overline{R}| > \square Q(N)\} < N^{-1} + \square^2 \quad , \tag{4}$$

This inequality, and the stronger version originally reported by Saw *et al*. (1984), represents a Tchebychev-type of inequality that is conditional on a sample mean $\overline{R}$ and a "pseudo-population variance" $Q(N)$ of $N$ realizations of a specified random variable, where $Q(N)$ = the product of $(N+1)/N$ and the corresponding $N$-sample variance, and where $\beta$ was assumed by Saw to be an arbitrary constant $\geq \sqrt{N/(N-1)}$. However, to derive the QR stated above, Woo (1991) assumed that $\beta = (N+1)\beta/Q(N)$ for arbitrary $\beta$. The latter assumption replaced Saw's *constant* $\beta$ by a function of the random variable $Q(N)$, thus rendering that function a *random variable*. The mathematical validity of Eq. (4) conditional on this replacement is not evident and was not proven by Woo (1991). Replacement of random estimators by corresponding sample statistics is often performed in statistics for the purpose of deriving new estimators. However, this type of substitution generally has probabilistic consequences that must be defined before the statistical utility of any such new estimator can be evaluated. The Woo (1991) QR moreover does not define a new estimator, but rather a rule asserting an absolute bound on error associated with sequential Monte Carlo estimates—a purpose to which the applicability of the underlying inequality was not claimed by Saw *et al*. (1984), has not since been demonstrated, and therefore remains questionable.

Besides its questionable mathematical basis, there is a fundamental practical problem associated with applying the Woo QR in the context of YMP-related uncertainty analysis. This problem is simply that the required minimum sample number $J_N$ can never be defined sufficiently reliably to ensure that future samples (e.g., any that might reflect as yet unsampled joint upper-tail input values that profoundly affect output, as discussed in Section 1.1) would not have the effect of contradicting the assumed $J_N$ definition conditional on any current sample size $N$.

### 1.3. A valid QR (VQR) may not be applicable to the TSPA

If some valid QR (VQR) along the lines of the Woo (1991) QR were correctly formulated and correctly applied to one or more TSPA submodels (e.g., the SZ model) in order to obtain random sample values of submodel output to use as inputs for Monte Carlo TSPA evaluation, a subsequent VQR application to evaluate the TSPA itself (i.e., a nested application of the VQR ) may or may not be valid. Certainly the validity of nested VQR applications would have to be established. This issue was not addressed

by Woo (1991). By definition, a VQR would characterize uncertainty in $\overline{R}$, the estimated mean (i.e., estimated expected) value of model output, as a function of $N$. If input variables to the model considered are themselves distributed and are (all, or for some variables) each sampled from a set of values the size of which set was determined by a previous VQR application, then the final VQR application may lack validity, particularly if the final model is a non-linear function of the distributed input variables. Only for a "general linear function" of random variables is the expected output value of that function equal to that same function applied to the expected values of the input variables. Thus, there is no general reason to assume that limiting uncertainty in the mean values of inputs will yield similar, or even predictable, limits on the uncertainty of the model output. Insofar as the TSPA model is non-linear with respect to some of its distributed inputs to which a valid VQR is applied, application of the same VQR to the TSPA model would not necessarily be valid, and the validity of applying any VQR to the TSPA in this situation would have to be established.

## 2. Strategic Discrete Probability Calculus (SDPC): An Appropriate, Feasible Way to Assess Reliability of Monte Carlo Estimates of SZ (or TSPA) Model Output

A practical problem that arises when trying to assess reliability of Monte Carlo estimates of SZ (or TSPA) model output concerns the scope or level of detail of the reliability assessment to be performed. Until now, evidently, this problem has been interpreted as being one of determining the appropriate Monte Carlo sample size, $n$, to use for the input-variable vector when conducting LHS to obtain model outputs. This size $n$ is just the number of input-variable vector realizations that are used to obtain $n$ corresponding model-output realizations. If the goal is to estimate the expected value of model output, then this problem is of little practical consequence so long as: (1) output-function evaluations are not expensive to obtain; (2) excessively heavy tails do not pertain to each input variable that alone, or interacting with one or more other input variables, has a substantial impact on the expected value of model output; and (3) the output model does not incorporate highly nonlinear interactions pertaining to the output effect of the input-variable values. In the context of YMP modeling, at least criterion number (1) above is violated for at least the SZ and TSPA models. Consequently, in this context, some fundamental limitations of (even LHS) Monte Carlo estimation must be considered.

Unless it can be ruled out *a priori* that input-variable values can interact strongly to affect output, then reasonable possibilities for such interaction must be assessed quantitatively if the reliability of estimated mean output is to be assessed with any confidence. This is difficult to do if LHS is relied on using a value of $n$ that is "small" relative to the number of potentially important interactive domains of the "parameter space" of the model being investigated. To see why, consider the possibility that values in the highest $100p\%$ tail of one or more of $v$ input variables may have a very large positive effect on mean model output, where each input variable is defined so as to have a nonnegative correlation with model output. Using LHS with a specified sample size $n$, the likelihood of obtaining a total of zero $k$-tuple upper-tail combinations is given by $p_o(n,p,k) = (1 - p^{k-1})^{np}$, where $k \leq v$ and where $np$ is here assumed to be a positive integer. Thus, for a model with only two variables suspected to be capable of substantial interaction ($k = 2$), a sample size of $n > 284$ would be required to have at least 95% confidence that the $n$ output realizations generated happen to contain *even a single* realization produced using an upper-tail value for each of the two variables of concern. For a model with three variables that might interact to substantially affect mean output ($k \geq 3$), a sample size of $n > 2{,}980$ would be required to have at least 95% confidence (or $n > 689$ would be required to have at least 50% confidence) that the $n$ output realizations contain ≥1 realization involving an upper-tail value for each of the three variables of concern. Thus, if interactions that substantially affect mean output cannot be ruled out *a priori*, fairly large sample sizes may be required even to search the relevant sample space so that potentially important areas of it are sampled *once*, let alone sampled reliably or systematically.

### 2.1. Strategic Discrete Probability Calculus (SDPC)

One alternative to uniform or Latin Hypercube Monte Carlo sampling is to perform Monte Carlo using an "importance sampling" strategy, which places greater-than-random emphasis on ranges of input-variable values, and/or on combinations of such value ranges, that are more likely to affect output. If samples are relatively expensive to generate (as in the YMP context), then inefficiency of the type discussed in Section 1.1 that is induced by Monte Carlo sampling can be avoided by dispensing with Monte Carlo sampling altogether as a means of output estimation and uncertainty analysis,

and by instead using Discrete Probability Calculus (DPC) (see Kaplan, 1981). DPC calculations are entirely analytic ones that involve combinations of fixed input-variable values and fixed corresponding probabilities.

The probability distribution function (pdf) of a stochastic input variable may be discrete (e.g., involving $v$ different values) or continuous. Any continuous or $v$-point discrete pdf may itself be approximated by a corresponding $m$-point discrete probability mass function (pmf), i.e., by an $m$-point approximating pmf (apmf), which apmf approaches an arbitrary degree of precision as $m \to \infty$ (for a continuous parent pdf) or as $m \to v$ with $m \leq v$ (for a discrete parent pmf). Of course, if the goal is to estimate the expected output value, general linear systems theory ensures that use of 1-point apmfs for each stochastic input variable (in this case, the expected value of each input) will yield the correct (exact) expected value of output, provided that output is modeled as a linear function of all the inputs. For nonlinear models, values of $m > 1$ must be used to obtain better-than-first-order estimates, the accuracy of which will increase by increasing the size of $m$. However, to the extent that the expected output value is affected primarily or substantially by relatively rare input-values or input-value combinations that yield relatively large output values, reasonably accurate estimates of expected output may still be obtained using values of $m$ as little as 2, provided that apmfs used are defined strategically so that potentially important combinations of upper-bound input values (in view of their respective likelihoods of occurrence) are represented in the DPC expression used to estimate the mean (and complete distribution) of model output.

An illustration of how 2-point apmfs may be used in the context of environmental health risk assessment was reported by Bogen (1995). This illustration showed how a 2-point apmf approach is most conveniently applied by summarizing each stochastic input variable by its expected value $\mu$ and by the expected value $\mu_p$ of the left-truncated pdf comprising the upper $100p\%$ tail of the parent pdf. By this approach, a skewed dichotomous pdf is created for each stochastic input variable. For example, assuming that $p = 0.10$, $\mu_p = 0.95$ for a uniformly distributed U(0,1) variable; see Bogen (1995) for a convenient expression to obtain $\mu_p$ for lognormally distributed variables. The apmfs involved using the approach of Bogen (1995) are defined to be "mean-preserving." Thus, if the output model is a linear function of model inputs, then the Bogen (1995) 2-

point approach ensures that the corresponding DPC-estimate of model output will always be exact.

If a 2-point apmf approach is used to perform DPC for a model that involves $v$ stochastic variables, then a total of $n = 2^v$ model-function evaluation steps are required to obtain the corresponding apmf of the model output, from which a corresponding (apmf-approximated) expected value may readily be calculated (as the ordinate-weighted sum of apmf abscissa values). For nonlinear models, the latter estimate of expected model output will not be exact, but confidence in the estimate arises from the fact that it (by definition) accounts for *all possible combinations* of upper-bound input values. This accounting for *all possible combinations* of upper-bound input values does not necessarily occur if LHS is used to obtain a Monte Carlo estimate of expected output, as noted above. As shown above, for a model involving three stochastic variables, a total of $n > 689$ model-function evaluations are needed to assure with >50% confidence that all possible combinations of upper-bound input values are each sampled at least once. Using DPC, that assurance is provided with 100% confidence using a total of only $n = 8$ model-function evaluations.

If model-function evaluation steps are expensive and $v$ is relatively large, then even a 2-point DPC approach to estimating model output and its uncertainty may not be feasible. In this case, there is no "perfect solution" to the problem of characterizing uncertainty in an estimate of expected model output. The relative merits of alternative approaches must be evaluated in terms of the extent to which they provide reasonable assurance that all potentially important areas of parameter-vector space are reflected in a corresponding estimated value of expected output. Considerations addressed above indicate that, in this regard, LHS may not provide as much reasonable assurance as a strategic implementation of 2-point-apmf-based DPC, hereafter referred to as "strategic DPC" (SDPC).

The SDPC approach recommended here is one in which a strategically truncated version of 2-point-apmf-based DPC is used to estimate expected model output. Truncation is accomplished either (1) by reducing the number $v$ of stochastic variables included in the model (as described below), or (2) by including in the DPC analysis only combinations involving $k$ or fewer of the $v$ stochastic variables included in the model.

By the first approach, SDPC implementation is facilitated by reducing $v$ to include only those stochastic inputs having distributional forms that are each known (or, from previous sensitivity studies, reasonably suspected) to have a substantial effect on mean output. Distributions pertaining to remaining variables should each be replaced by their corresponding expected values (i.e., $v <$ the original total number of stochastic parameters).

By the second approach, DPC analysis addresses all $v$ input variables, but includes only combinations involving $k$ or fewer of the $v$ stochastic variables included in the model. For reasonably small values of $p$ defined above (e.g., $p = 0.10$), the likelihood $p^k$ of each $k$-tuple upper-bound combination clearly becomes small quite rapidly as $k$ increases. This in turn means that, even for moderately large $k$, if the contribution of one or more particular $k$-tuple upper-bound combinations to the expected output value happen to be quite large (relative to a conditional expected value calculated without these combinations), such contributions may nevertheless have minimal impact on actual expected model output due to their very small likelihoods of occurrence. Values for $p$ and $k$ selected thus reflect a tradeoff between potentially important effects on estimated mean output that depend on the rarity ($p$) of the range of upper-bound values that most affect mean output for each variable, and the likely magnitude of $k$-tuple upper-bound interactive effects on mean output. Of course, if setting $k<v$ is used as the method of DPC-problem truncation, then the likelihoods $p_j$ of each $j$th combination of (non-upper-bound and/or upper-bound) input values considered will sum to a value $P \leqslant 1$, and the corresponding output apmf must therefore be normalized by dividing all $p_j$ values by $P$.

The SDPC approach outlined above has the important advantage that *all possible k-tuple combinations of upper-bound variable values* will in fact be incorporated into each SDPC-based estimate of expected model output. In contrast, a LHS approach offers no such assurance for problems involving relatively large $v$ when $n$ cannot be very large because model-output function evaluations are expensive. Variations in SDPC quantities (namely, $p$, $k$ and $v$ defined above) may be used to explore the sensitivity of SDPC-based estimates, relative to the estimate obtained using baseline quantities such as $p = 0.10$, $k = 2$, and $v =$ the original total number of stochastic parameters. To examine the impact of model nonlinearity, it is also easy to compare estimates of mean

output using 2-point SDPC analysis to corresponding first-order estimates obtained using the 1-point approach (i.e., using expected values for each stochastic input variable, e.g., as indicated for SZ outputs in Figure 11 on p. 43 of BSC, 2001a).

**2.2. Application of SDPC to assess reliability if estimated mean SZ/TSPA output**

The SZ model has $v = 29$ stochastic input variables (BSC, 2001b), which the SZ model uses to calculate output radionuclide concentrations at a 20-Km site fenceline from corresponding unit-valued radionuclide concentrations assumed to be input from the unsaturated zone (UZ) (TRW, 2000). Even using 2-point apmfs for each input variable, a standard DPC analysis would thus require $n = 2^{29} = \sim 5.4 \times 10^8$ model-function evaluations, which is too large to be practical in the case of SZ (or TSPA) model-output analysis. Using truncation method 1 referred to in section 2.1 above, the scope of the problem might reasonably be reduced strategically to require $n' < n$ output evaluations by focusing only on a subset of $v' < v$ input variables known to have the greatest impact on expected model output, based on preliminary sensitivity analyses (such as a rank-regression analysis performed on output from a previous LHS Monte Carlo analysis). For example, six SZ parameters (see Table 15 on pp. 65-68 of BSC, 2001b) that each appear to have a particularly large univariate impact on expected SZ output, relative to that of the other input parameters, are: Groundwater specific discharge (GWSPD), Flowing interval porosity (FISVO), Diffusion coefficient (DCVO), Sorption $K_d$ coefficient for Np (KDNPAL), Horizontal anisotropy (HAVO), and Colloid retardation factor in the alluvium (CORAL) (Stephanie Kuzio, YMP, Sandia; personal communication, 26 March 2003). Additional parameters might also be considered as feasible. SDPC results might thus be compared based on $v' = 6$, 7 and 8, corresponding to a total of $n' = 64$, 128 and 256 output evaluations, respectively.

Alternatively, using truncation method 2 referred to in section 2.1 above, the scope of the SZ problem might reasonably be reduced strategically to address only $k$-tuple combinations of input parameters that each are dichotomized using 2-point apmfs, where $k < 29$. If $k = 2$, then a total of $n = \binom{29}{2} = 406$ output-function evaluations would be required (which might be feasible); for $k = 3$, a total of $n = 3{,}645$ evaluations would be needed (which would not likely be feasible). A combination of both methods might

also be employed, e.g., to obtain SDPC estimates of expected output for all $k$-tuple combinations of $v'$ 2-point approximations of the most important parameters, where $1 \leq k \leq 3$. By this "mixed-truncation" method, using $v' = 8$, 10 or 12 corresponds to a requirement for a total of $n' = 92$, 175 or 298 output evaluations, respectively, whereas if $1 \leq k \leq 2$, then using $v' = 20$, 25, or 29 corresponds to a requirement for a total of $n' = 210$, 325 or 435 output evaluations, respectively. The "mixed-truncation" method thus has the advantage of providing a nonstochastic estimation of expected model output that accounts for *all possible* upper-bound combinations involving up to a total of Max($k$) among $v'$ stochastic input variables. A demonstration that estimates obtained using increasing values of Max($k$) are similar in magnitude would support the conclusion that these estimates are reliable. Importantly, such an assessment of reliability would involve the most reasonably comprehensive systematic exploration of the impact of potentially important parameter interactions on estimated model output that is possible, conditional on the greatest feasible number of output function evaluations. A similar claim cannot be made if a LHS Monte Carlo approach is used to estimate expected model output.

Expected SZ-model outputs have been presented in terms of relative-mass-flux breakthrough curves for radionuclide or radionuclide class as functions of time (see Figure 11 on p. 43 of BSC, 2001a). Reliability assessment of expected SZ-model output (by whatever method) should also include analyses performed using an SZ-output metric that corresponds directly to TSPA output—that is, to radionuclide-specific contributions to the expected value, E$D$ (where E is the expectation operator), of maximum 1-year committed dose $D$ to a hypothetical resident at the 20-Km site fenceline, conditional on the corresponding concentration input $I_i(t)$ for that $i$th radionuclide from the Unsaturated Zone (UZ) as a function of time $t$ assuming that $I_i(t) \equiv 1$ for all $t$. The latter allows the reliability assessment to be interpreted more easily in terms of consequences for TSPA modeling than an assessment that focuses on the reliability of particular estimated SZ-specific outputs, some of which will have little impact on the final TSPA output of ultimate concern. As noted above, the SZ model generates radionuclide-specific concentrations $C_i(t)$ at time $t$ conditional on corresponding unit input UZ concentrations. Each $C_i(t)$ value is linearly related to a corresponding committed-dose contribution $D_i(t)$ (conditional on unit-concentration UZ

inputs) by the equation: $D_i(t) = F_i C_i(t)$, where $F_i$ is the Biosphere Dose Conversion Factor (BDCF) for the corresponding $i$th radionuclide. Uncertainty distributions (and very reliable estimates of the expected values $EF_i$) for the factors $F_i$ can be generated readily from the Biosphere model, and are currently available. Translation of SZ outputs into corresponding TSPA-related outputs is facilitated by the fact that $ED_i(t) = EF_i\, EC_i(t)$, such that $ED = \mathrm{Max}(\sum_i^g EF_i\, ED_i(t) EI_i(t))$, where $g$ is the number of radionuclides and Max is taken with respect to all times $t$, and where calculation of $ED$ would require estimates of $I_i(t)$ from the UZ model (and so might be beyond the scope of SZ-output reliability assessment, but would be required to assess reliability of estimated mean TSPA output). Note that the relations defining $ED_i(t)$ and $ED$ are not affected by the covariance structure of $C_i(t)$ for all $i$. To reduce calculations for the purpose of assessing reliability of SZ-generated $ED_i(t)$ estimates, times $t$ might be evaluated only at 50- or 100-year increments.

**Acknowledgments**
The author is grateful for the initial request for this work and for comments provided by Anthony Smith (Biosphere Group Leader) of the Yucca Mountain Project program at Bechtel SAIC Company (BSC), Las Vegas, NV, by Stephanie P. Kuzio (Saturated Zone Group Co-Leader) of the Yucca Mountain Project at Sandia National Laboratories, Albuquerque, NM, by Jane (Xiaoyan) Ji (Systems and Decision Sciences, LLNL), and by Jeff Daniels (Risk Sciences Group Leader, LLNL). I am also grateful for comments provided to me in March 2003 by Professor Mark C.K. Yang concerning the proper interpretation and application of results reported in the paper by Saw *et al*. (1984). This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

**References**


Bechtel SAIC Company (BSC). 2001a. Input and Results of the Base Case Saturated Zone Flow and Transport Model for TSPA (Unit breakthrough curves from the SZ site-scale flow and transport model used in the TSPA-SR simulations). ANL-NBS-HS-000030 Rev. 00 ICN 01 (September 2001).

Bechtel SAIC Company (BSC). 2001b. Uncertainty Distribution for Stochastic Parameters (Development of SZ stochastic and constant parameters that are input to the SZ site-scale model for the TSPA-SR). ANL-NBS-MD-000011 REV 00 ICN 01 (October 2001).

Bellman, R. 1961. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, Princeton, NJ.

Bogen, KT. 1995. Methods to approximate joint uncertainty and variability in risk. *Risk Anal*ysis **15**, 411-19.

Feller, W. 1970. *An Introduction to Probability Theory and it Applications, Vol. I*, 3rd ed. (revised printing). John Wiley & Sons, New York, NY, p. 106.

Helton JC. 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive-waste disposal. *Reliability Engineering & System Safety* **42**, 327-67.

Helton JC, Davis FJ. 2002. *Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems*. SAND Report SAND2001-0417. Sandia National Laboratories, Albuquerque, NM.

Johnson, NL, and S Kotz. 1977. *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory*. John Wiley & Sons, New York, NY, p. 158.

Kaplan, S. 1981. On the method of discrete probability distributions in risk and reliability calculations—Application to seismic risk assessment. *Risk Anal*ysis **1**, 189-96.

Saw JG, Yang MCK, Tse CM. 1984. Tschebyschev inequality with estimated mean and variance. *American Statistician* **38**, 130-32.

TRW Environmental Safety Systems Inc. (TRW). 2000. Saturated Zone Flow and Transport Process Model Report. TDR-NBS-HS-000001 REV 00 ICN 02. Report prepared for the USDOE Yucca Mountain Site Characterization Office, PO Box 30307, North Las Vegas, NV (October 2002).

Woo, G. 1991. A quitting rule for Monte Carlo Simulation of extreme risks. *Reliability Engineering and System Safety* **31,** 179-89.